High Performance Computing OpenStack Options

September 22, 2015

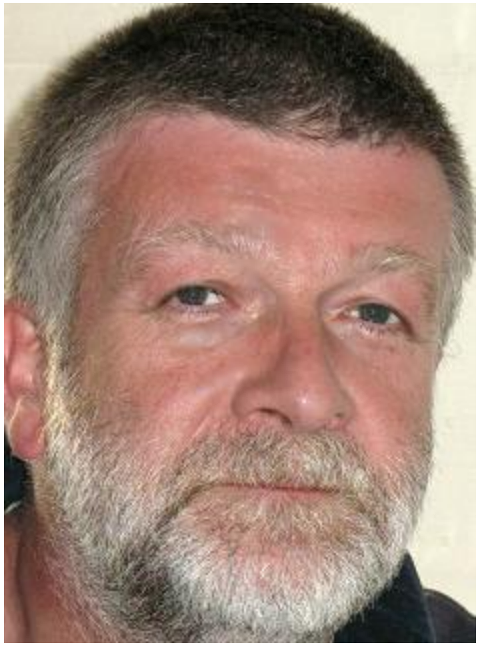
SNIA™

Cloud Storage Initiative

Today's Presenters

Glyn Bowden, SNIA Cloud Storage Initiative Board HP Helion Professional Services

Alex McDonald, SNIA Cloud Storage Initiative Chair - NetApp



SNIA Legal Notice

- The material contained in this tutorial is copyrighted by the SNIA unless otherwise noted.
- Member companies and individual members may use this material in presentations and literature under the following conditions:
  - Any slide or slides used must be reproduced in their entirety without modification
  - The SNIA must be acknowledged as the source of any material used in the body of any document containing material from these presentations.
- This presentation is a project of the SNIA Education Committee.
- Neither the author nor the presenter is an attorney and nothing in this presentation is intended to be, or should be construed as legal advice or an opinion of counsel. If you need legal advice or a legal opinion please contact your attorney.
- The information presented herein represents the author's personal opinion and current understanding of the relevant issues involved. The author, the presenter, and the SNIA do not assume any responsibility or liability for damages arising out of any reliance on or use of this information.

  NO WARRANTIES, EXPRESS OR IMPLIED. USE AT YOUR OWN RISK.

Abstract

Organisations are beginning to look to OpenStack to provide framework and tenancy controls around HPC workloads. The greatest gain in the multi-tenancy model is also the greatest challenge for storage; how to provide, reliable, high performance storage that is adequately segregated for the workloads in a cloud environment.

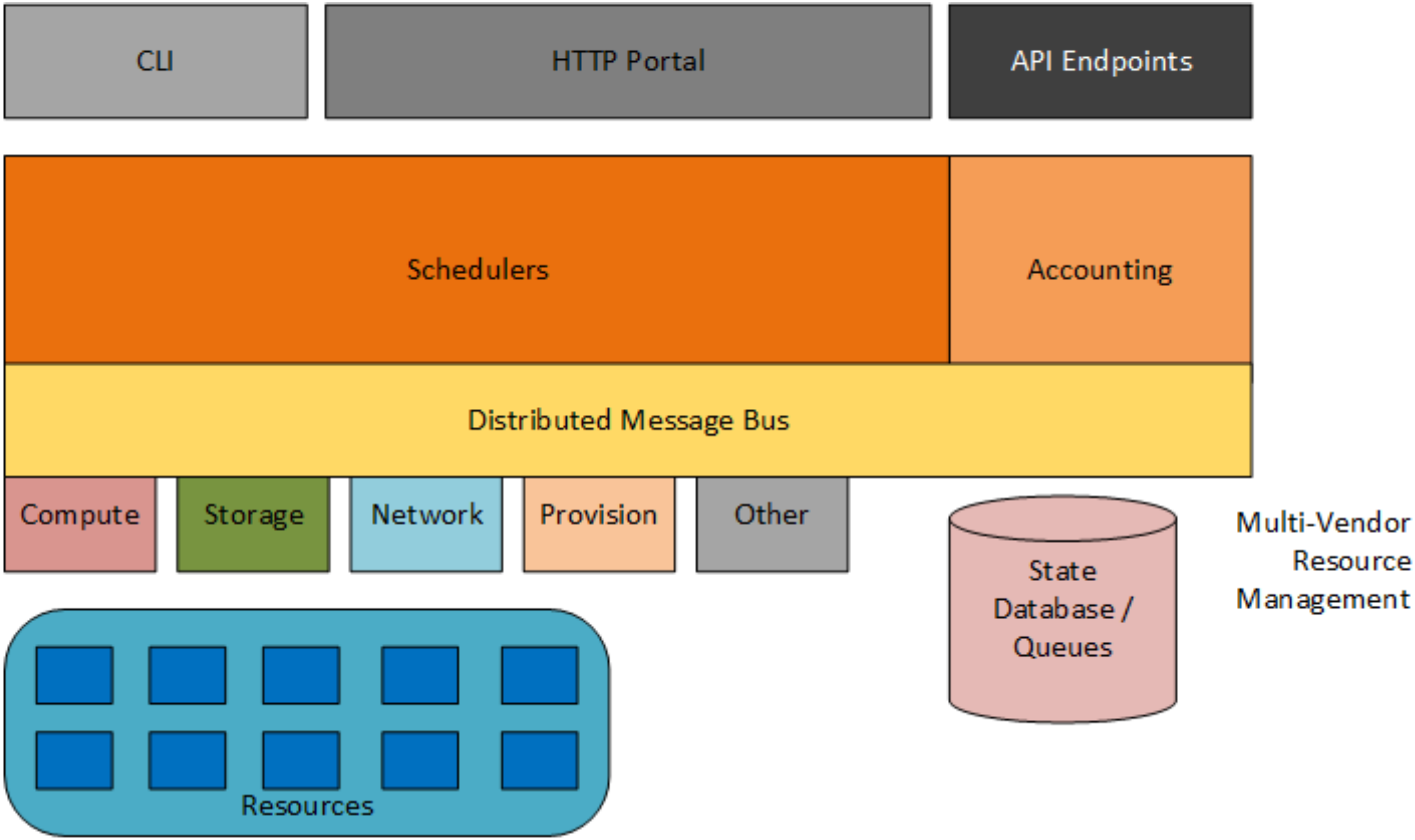
This presentation looks at the options available within OpenStack and the Cloud Storage community.

Agenda

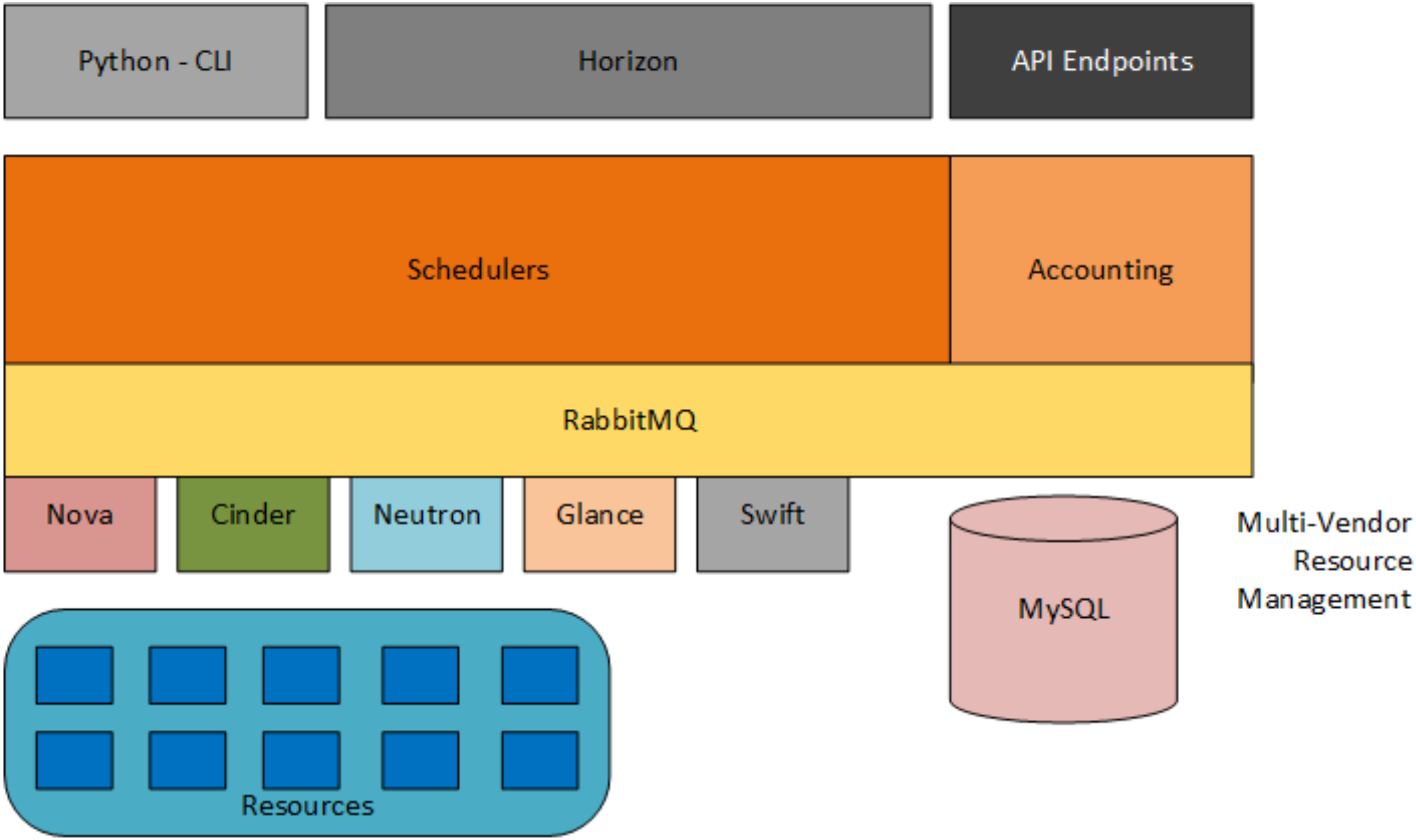
- HPC vs OpenStack
- What is “High Performance Compute”
- Specific Challenges
- Storage Options
  - Block
  - Object
  - File
- Example Scenario
- Summary

HPC VS OPENSTACK

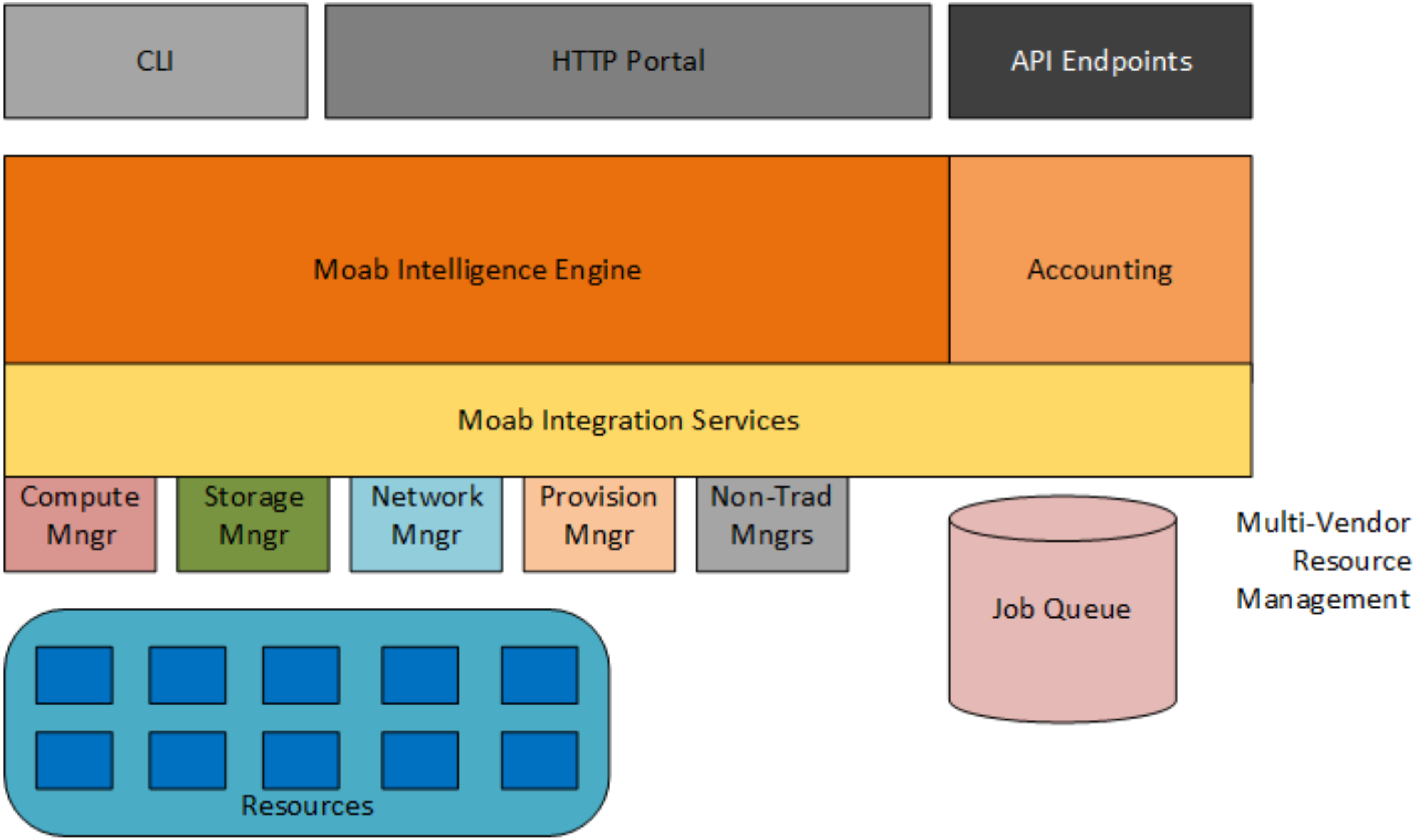
The logical architecture



..of OpenStack



..and MOAB HPC Suite



HPC and OpenStack – Opposing Forces

- Cloud
  - Share Everything
  - Generic Workloads
  - Loosely Coupled
  - Many small workloads

HPC and OpenStack – Opposing Forces

- Cloud
  - Share Everything
  - Generic Workloads
  - Loosely Coupled
  - Many small workloads

- HPC
  - Share Nothing
  - Specific, Niche Workloads
  - Tightly Coupled (RDMA)
  - Few Large Distributed Workload

But the same…

- Cloud
  - Highly Distributed
  - Large Storage Pools
  - Resource Management Key
  - Performance Management

- HPC
  - Highly Distributed
  - Large Storage Pools
  - Resource Management Key
  - Performance Management

SNIA™
Cloud Storage Initiative

WHAT IS HIGH PERFORMANCE COMPUTING?

Background on HPC

- Two Major Types of HPC
  - **Analytics**
  - Big Data Sets
  - Simple Operations repeated many times
  - Aggregation of results
  - **Computationally Intensive**
  - Smaller Data Sets
  - Very complex algorithms that need to be broken down
  - Sequential processing and summary
  - Often Latency Sensitive (RDMA, Lustre) or Bandwidth Sensitive (High Volume Filesystems, Large Files)

Background on HPC

- Two types of Computational HPC
  - **Batch Processing**
  - Loosely coupled
  - Embarrassingly Parallel
  - Limited / No shared resources during jobs
  - **Realtime / Grid Computing**
  - Tightly Coupled
  - Requires High Performance Networking for Remote Direct Memory Access (RDMA)
  - Usually a high performance shared file system is required
  - CPU and Memory Architecture more critical than batch

Why HPC and OpenStack

THE CHALLENGES

The Challenges

- Resource Management
  - HPC clusters have always been very good at managing their own resources
  - Challenge comes when security and multi-tenancy is required
- Multi-Tenancy Drivers
  - Genome Research driving separation
  - Human data cannot be shared beyond proposed use
  - Projects use flexible resources but on a fixed hardware platform

Storage Options

EPHEMERAL STORAGE

Ephemeral Storage

- What is it?
  - Persists only as long as the VM exists
  - Usually located locally on the compute server
- HPC Use Cases?
  - User scratch space
  - Work scratch space
  - Operating Environment

Storage Options

BLOCK STORAGE

Block Storage

- What is it?
  - Persistent, non-shared block storage
  - Can be provided by many sources, SAN Arrays, Local Disk etc.
- HPC Use Cases?
  - Supporting Databases
  - High performance scratch space
- OpenStack Project is CINDER
  - A Large Disk Array attached by Fibre Channel SAN to all of the compute nodes
  - OpenStack uses Cinder drivers to create, mount and protect LUNs for the guests.
  - Guest is responsible for creating a file system on those LUNs
  - Not shared with other guests

Storage Options

OBJECT STORAGE

Object Storage

- What is it?
  - Persistent, scalable storage pools
  - Access using a REST based API
  - Not bound to an individual Guest
- HPC Use Cases?
  - Centralised Data Lakes
  - Archives / Backups of source data
- OpenStack Project is Swift
  - Usually uses large pools of local disk attached directly to the object servers
  - Uses metadata to index the data and locate object blocks from unique identifiers
  - Lots of plugins for the various analytics engines that are expanding the adoption of object as a centralised data lake

Storage Options

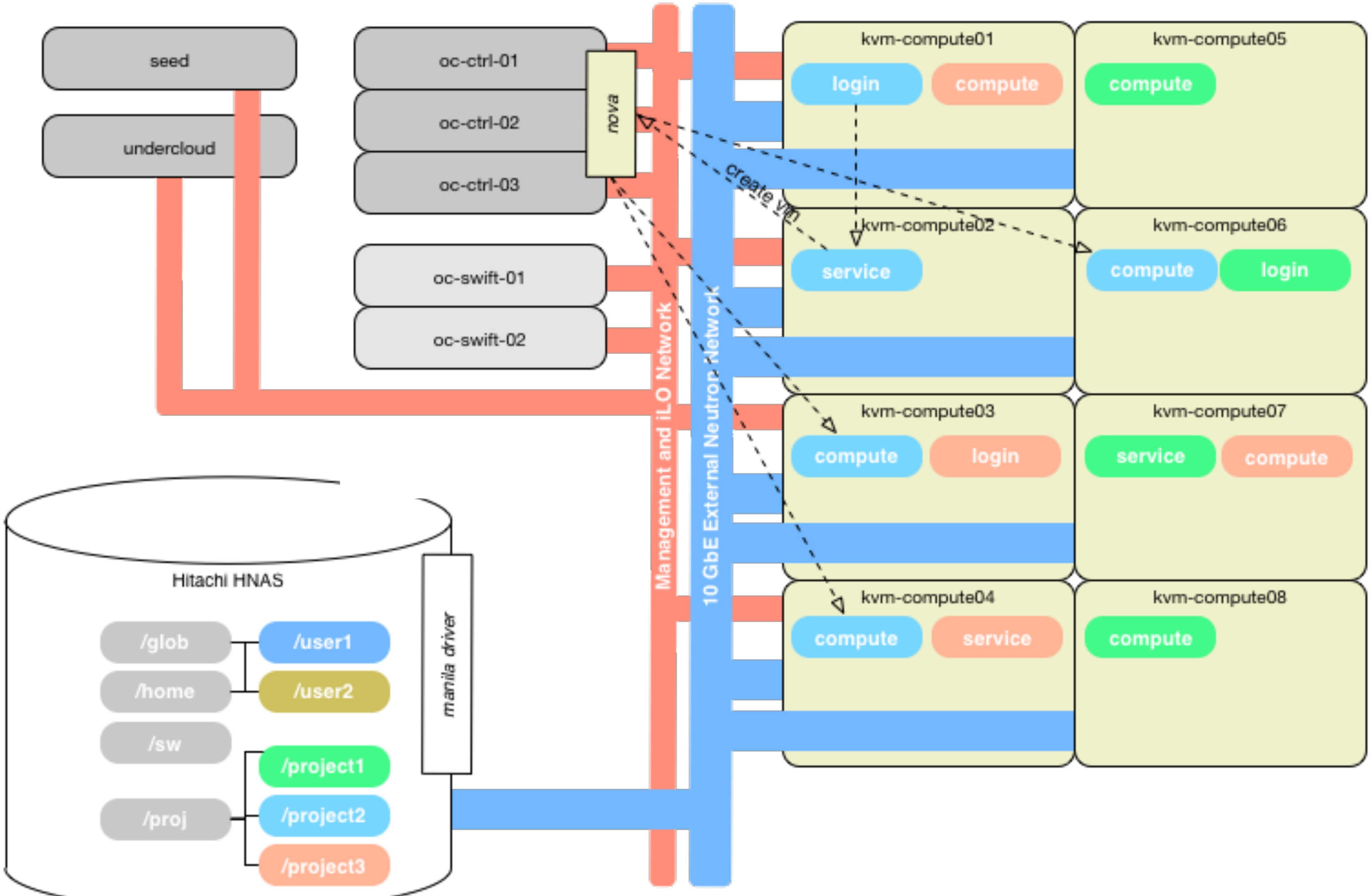
FILE STORAGE

File Storage

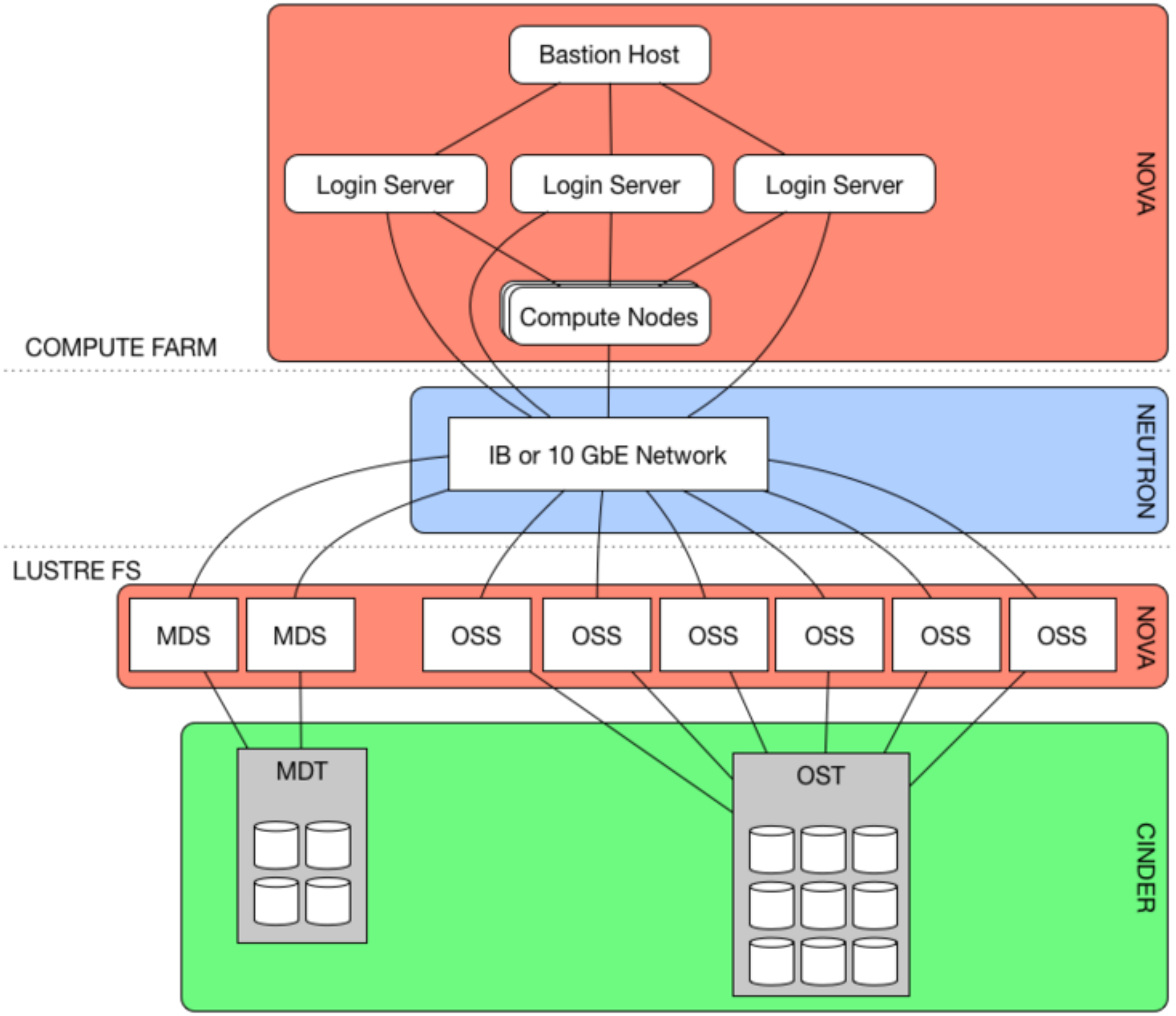
- What is it?
  - Shared, persistent storage
  - Uses standard POSIX file system methods to access data
- HPC Use Cases?
  - User Home Directories
  - Shared Project Data
  - Scale out file systems!
- OpenStack Project is Manila
  - Manage the creation of storage pools on the provider service
  - Create the shares and apply the correct permissions
  - Mount those shares within the guests that need them
  - Can be NFS or CIFS based today
  - Plugin driven

AN EXAMPLE SCENARIO

Storage Use Case

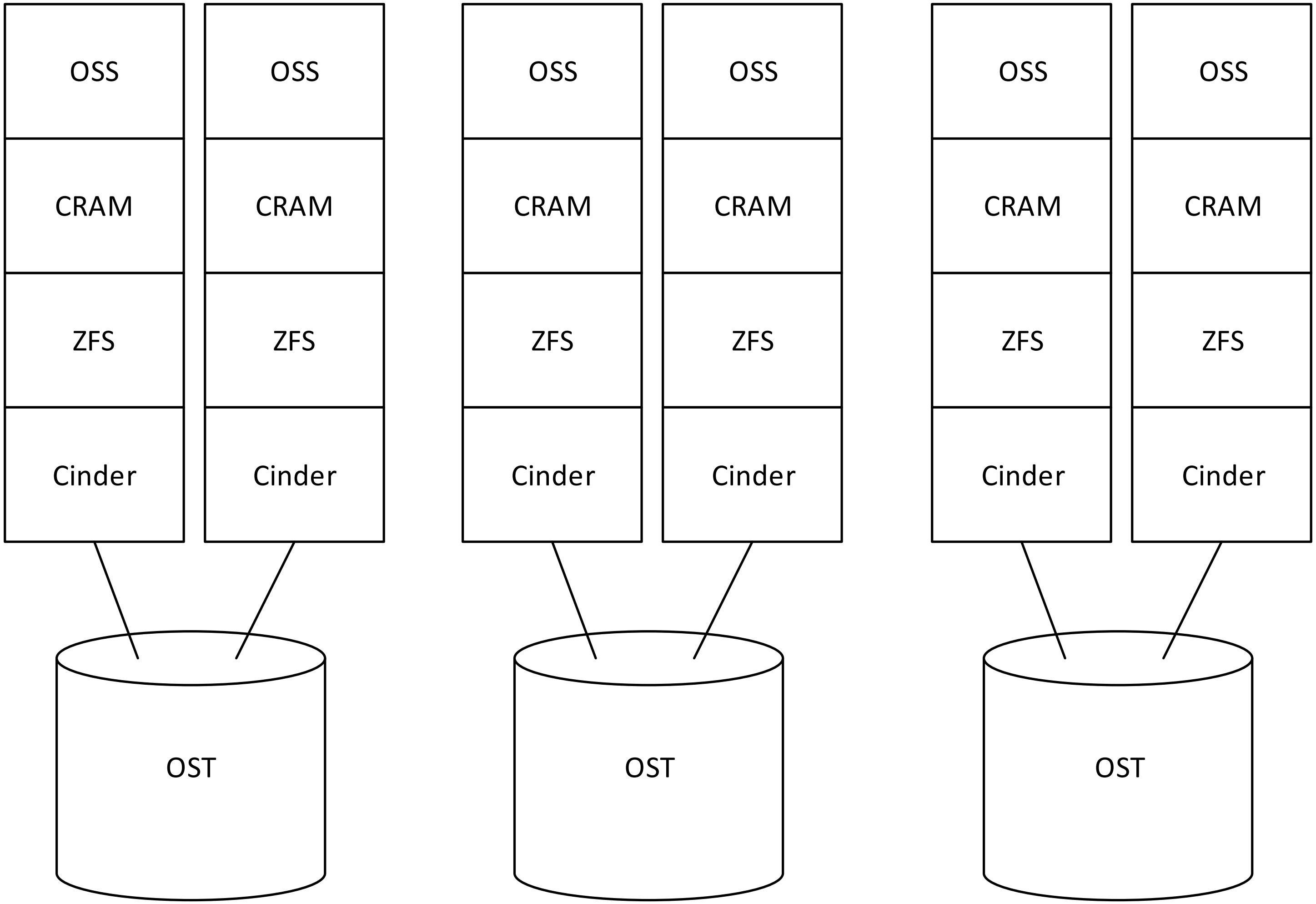


OpenStack and Lustre Mapping



Lustre as a Service Stack

SNIA
Cloud Storage Initiative



Lustre Components

SNIA™
Cloud Storage Initiative

- Massively Parallel Filesystem made up of key components…
  - MGS – Management Server
  - MGT – Management Target
  - MDS – Meta Data Server
  - MDT – Meta Data Target
  - OSS – Object Storage Server
  - OST – Object Storage Target
- 1 File can be spread over up to 2000 objects
- With ldiskfs, each of those each object can be up to 16 TB
- That's 31.25 PB (Yes PETA bytes) for a single file using ldiskfs
- Up to 4 Billion files per MDT
- Up to 4096 MDTs!

What about ZFS? Why?

- Lustre has limited data protection.
  - RAID 0
  - Protection from Physical Infrastructure
- Scale Out – Easy, Scale UP – Hard
- ZFS has healing, snapshots (not necessarily a good idea here) and scale up!
- ZFS Cache Pools for Meta-Data or even data sets, huge acceleration potential
- Scale…

Lustre + ZFS File Limits

	LDISKFS	ZFS
Object Size	16 TB	256 PB
Maximum File Size	21.25 PB	8 EB (2^63)
Max Files per MDT	4 Billion	4 Billion
Max MDTs	4096	4096

Work in Progress

- Lustre can be for high bandwidth and low latency
- Low latency challenging in virtual environment
- High Bandwidth, easier (not simple though)
- Use OpenStack tools to provision Lustre Components
- Build small scale, segregated clusters for multi-tenancy
- Export via NFS with Manila on private networks
- Include ZFS and Compression
- Use Cinder Block for the shared storage element

SUMMARY

Summary

- Initial interest of HPC on OpenStack is being driven by tenancy requirements
- Managing flexible HPC resources has been tricky, OpenStack makes that easier for HPCaaS
- Many areas are needed to work well together for success, OpenStack Community beginning to address that as we have seen.
- HPC on OS is a reality and many are pushing the boundaries and committing back to the community.

After This Webcast

SNIA Cloud Storage Initiative

- This webcast and a copy of the slides will be posted to the SNIA-CSI website and available on-demand
  - http://www.snia.org/forum/csi/knowledge/webcasts
- A full Q&A from this webcast, including answers to questions we couldn't get to today, will be posted to the SNIA Cloud blog
  - http://www.sniacloud.com/
- Follow us on Twitter @SNIACloud
- Upcoming SNIA Webcast: OpenStack Manila – Oct. 7th
  - https://www.brighttalk.com/webcast/663/173013
- Google Groups:
  - http://groups.google.com/group/snia-cloud

Conclusion

Questions

Conclusion

SNIA™
Cloud Storage Initiative

Thank You